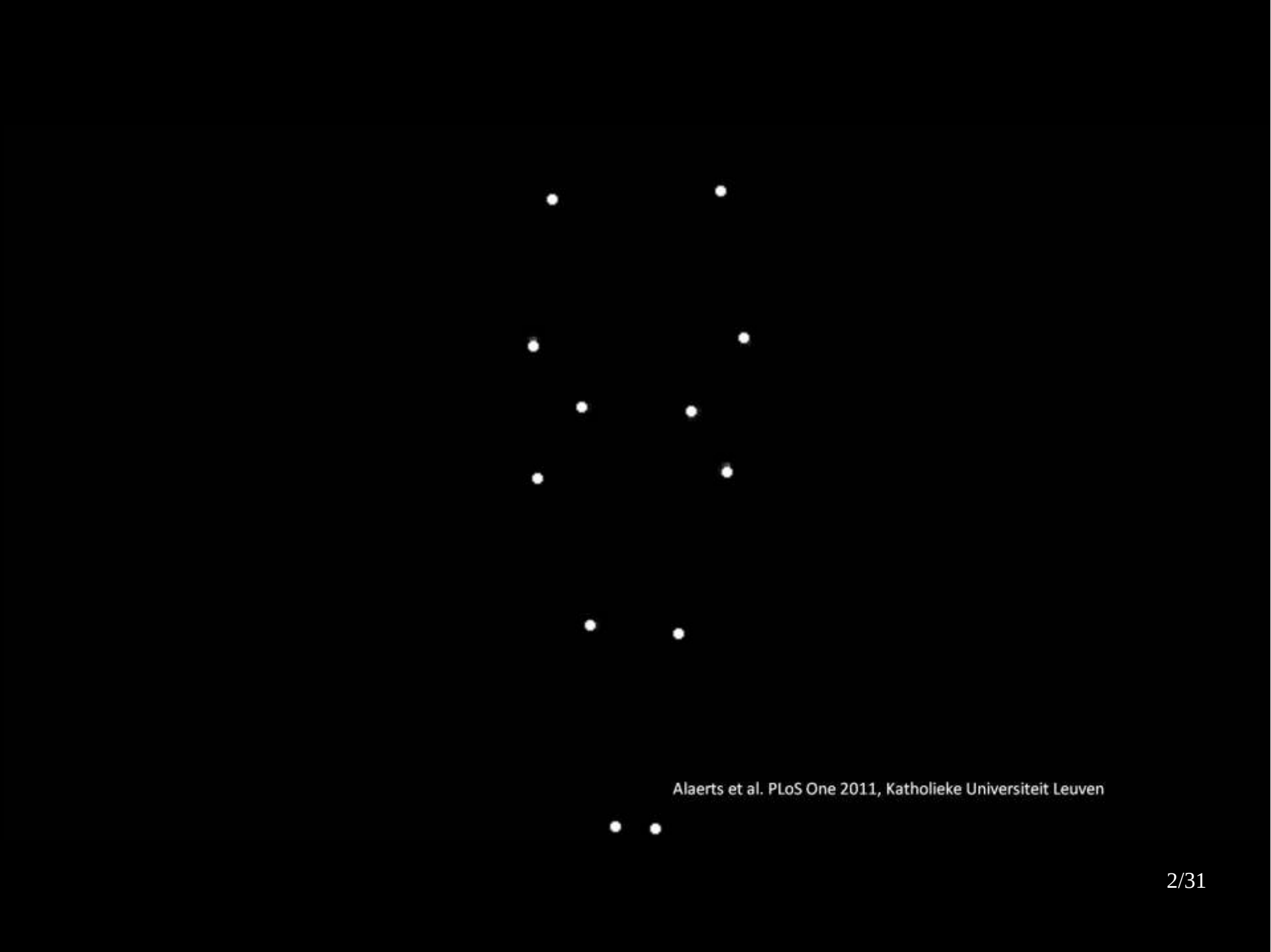
Perception: Structure and Association

Justus Piater
Intelligent and Interactive Systems
University of Innsbruck
Austria


Alaerts et al. PLoS One 2011, Katholieke Universiteit Leuven

Associative Models

- Perception evokes interpretation
  - via learned association between models and percepts.
- Learning reactive and forward models is building associations.
  - I perceive this – then I do that.
  - I do this – then I perceive that.

Association is not...

- bottom-up reconstruction
  - structure from motion
  - physical models
- just labels
  - object, category, …
- bottom-up segmentation
- top-down segmentation
- static

classical computer vision

Association...

- learns contingencies
- evokes models from observations
- expects observations from models
- is relevant to action
- is dynamic

Structure

- geometry
  - pose
  - kinematics
  - ...
- functional relations

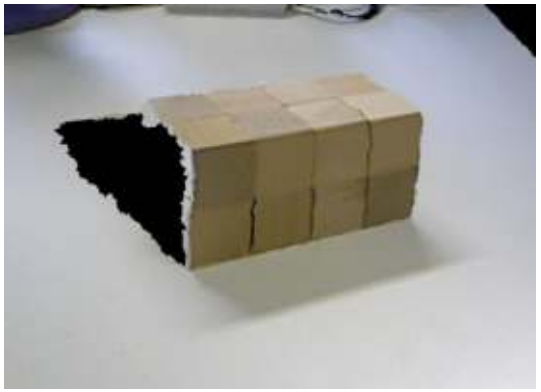
Structure and Association go hand in hand.

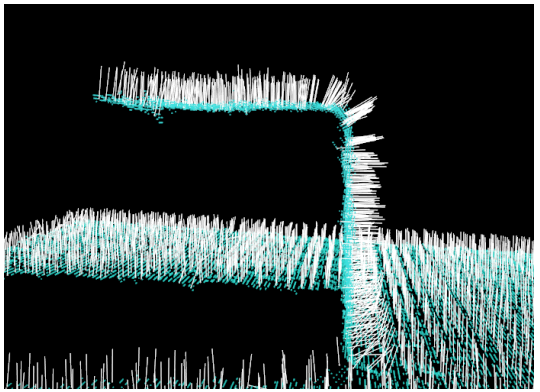
3D Object Partitioning

3D structural building blocks (inspired by perceptual psychology)

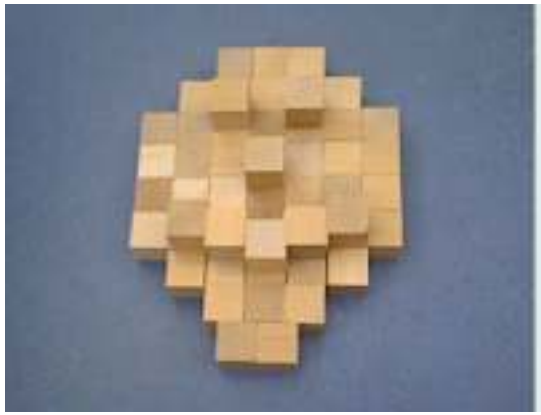
Hardware:
RGB-D(epth) cameras
(„3D-camera“)

Data to process:
Colored 3D point clouds
(x,y,z, r,g,b)

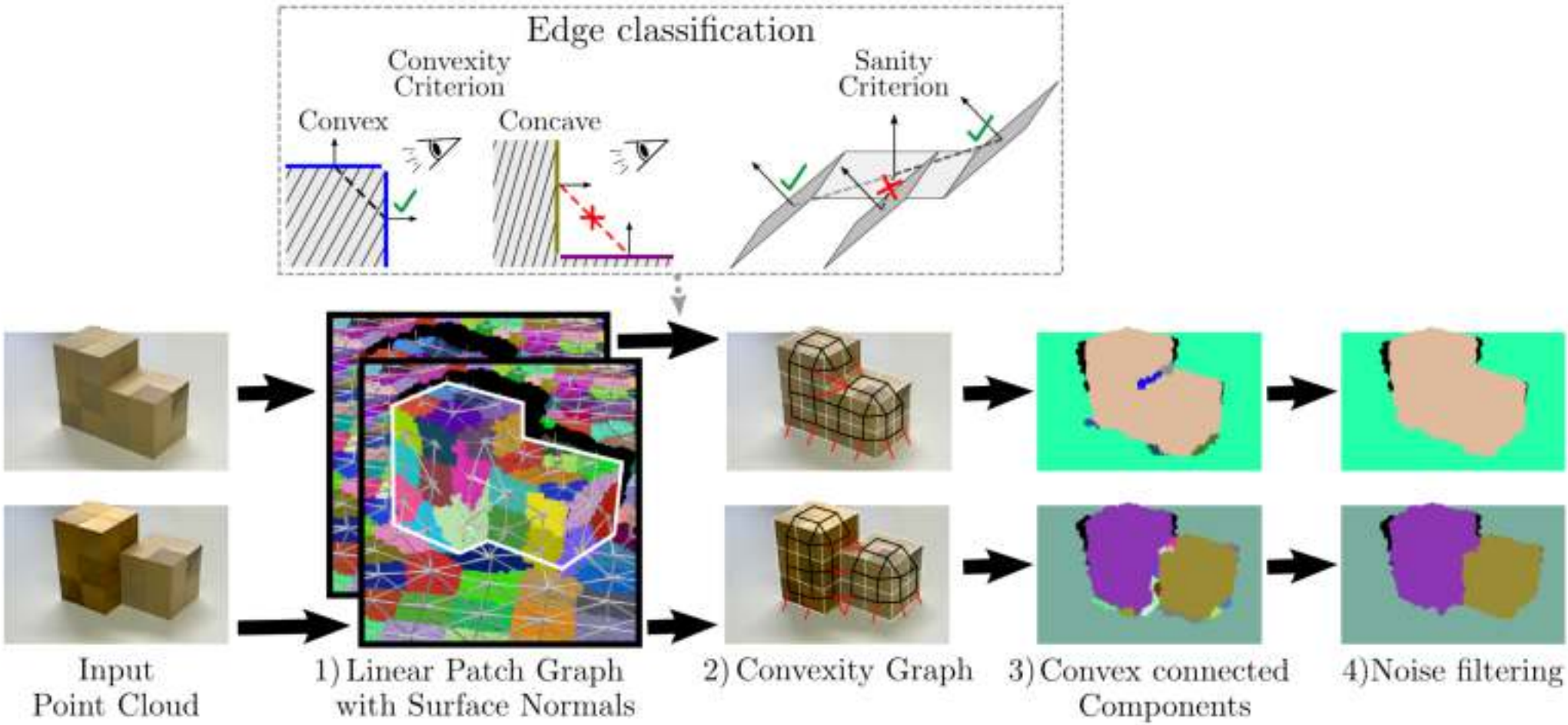




Goal:
Partitioning of data into
”objects/parts”



3D Object Partitioning: Algorithm

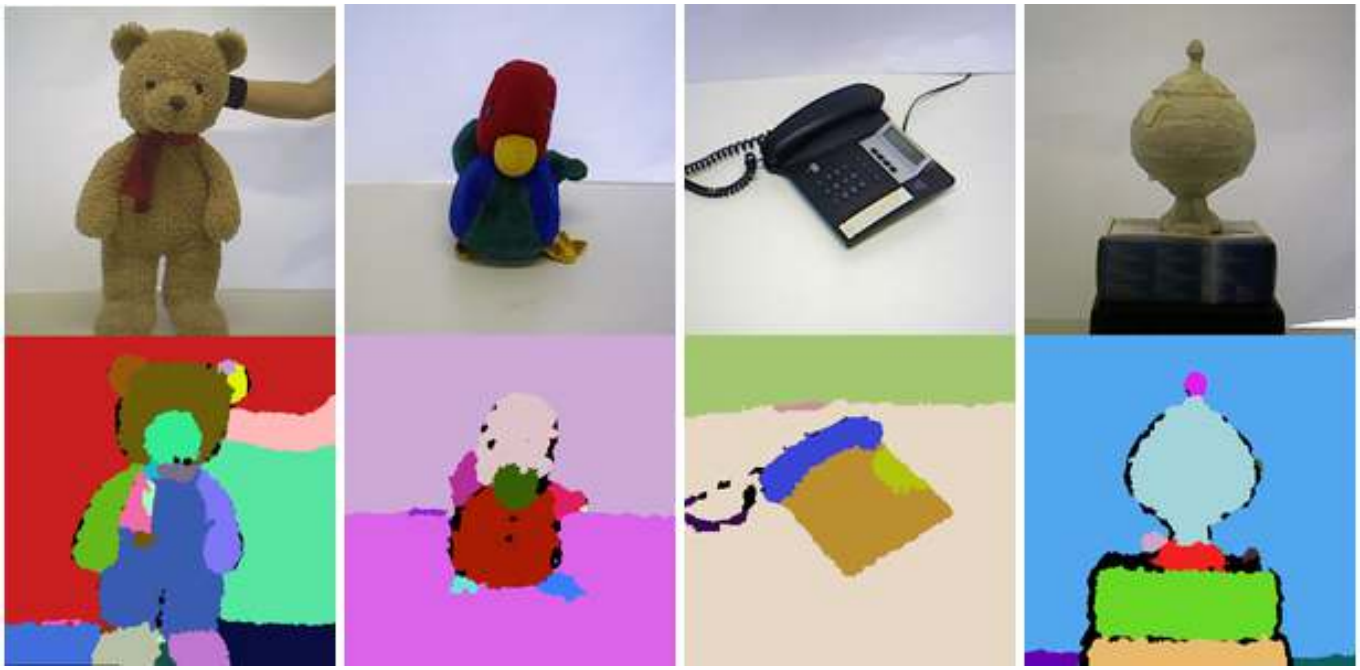


- Basic Algorithm: S. C. Stein, F. Wörgötter, M. Schoeler, J. Papon, T. Kulvicius, "Convexity based object partitioning for robot applications", ICRA 2014
- Extension: S. C. Stein, M. Schoeler, J. Papon, F. Wörgötter, "Object Partitioning using Local Convexity", CVPR 2014

Example results

“Simple” objects
(OSD dataset)

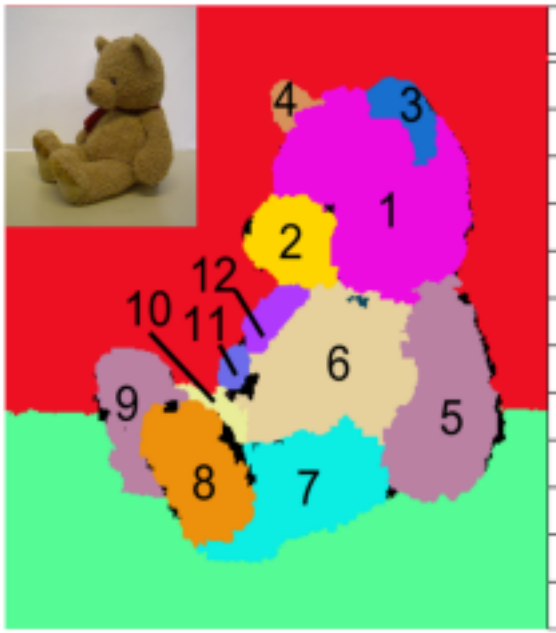
“Part”
Segmentation



Example results

Complex realistic scenes (NYU dataset)

Partitioning of high quality / density data

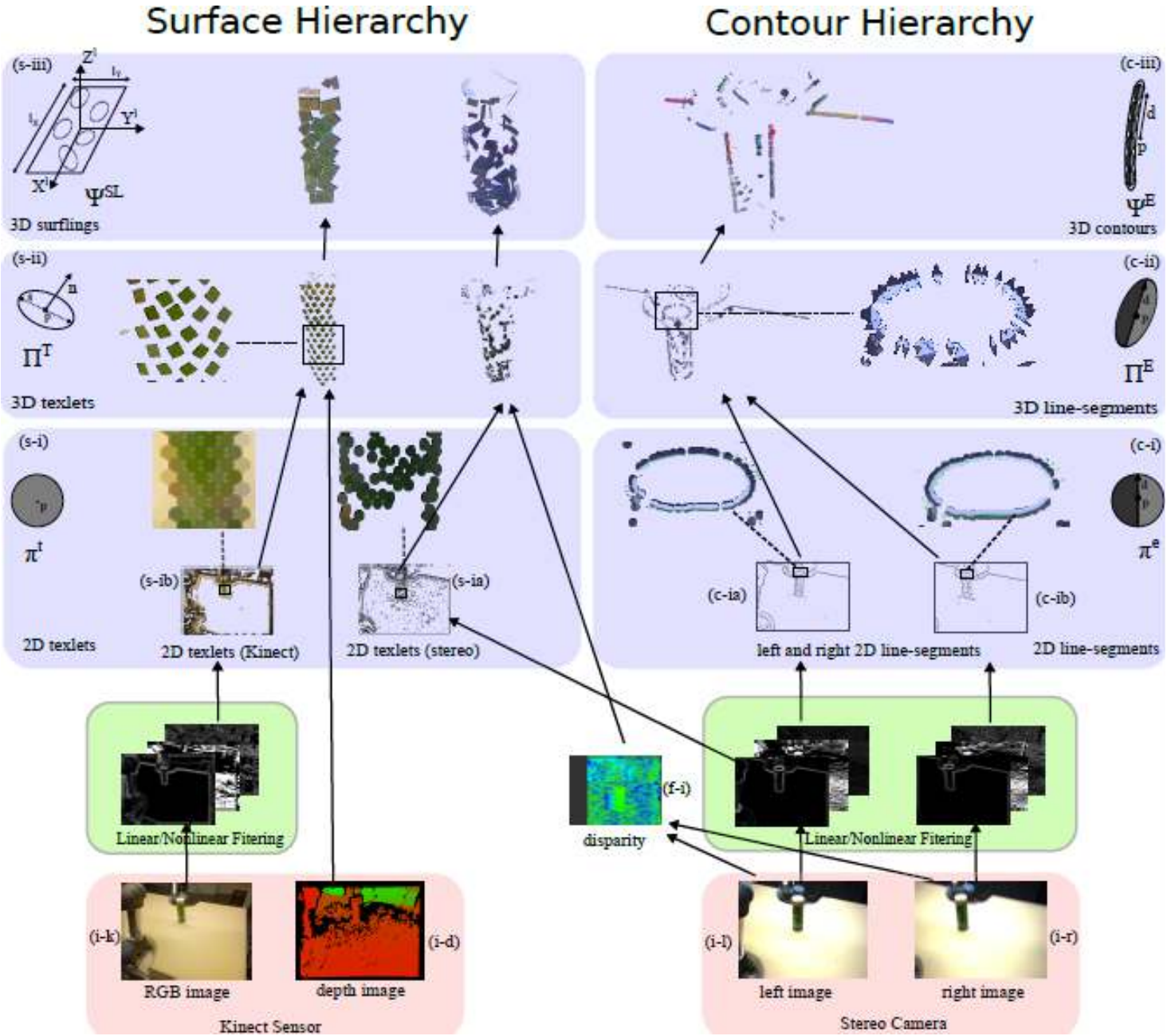


ID	name
1	head
2	snout
3	left ear
4	right ear
5	left arm
6	belly
7	left leg
8	left foot
9	right foot
10	right leg
11	right arm
12	tie

Early Cognitive Vision

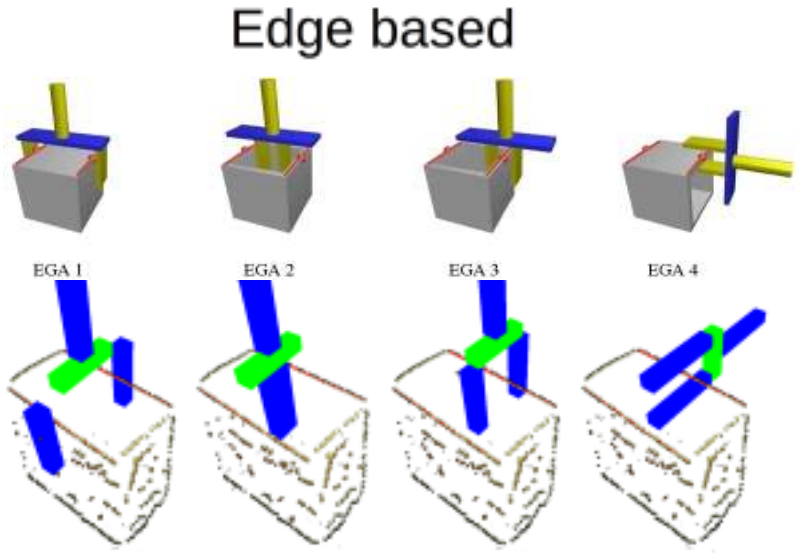
Hierarchies of semantic structure

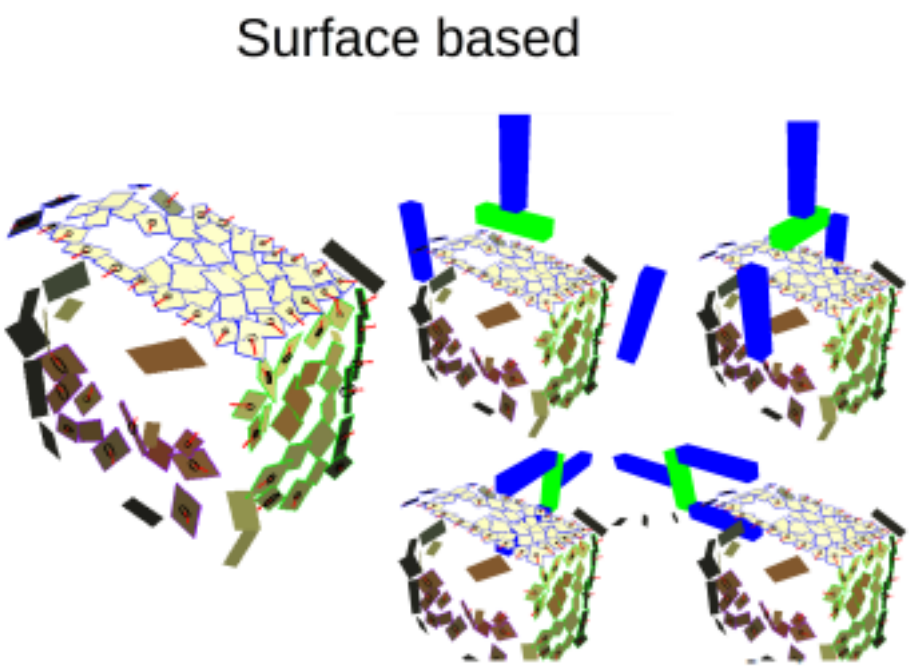
(inspired by neuro-physiology)

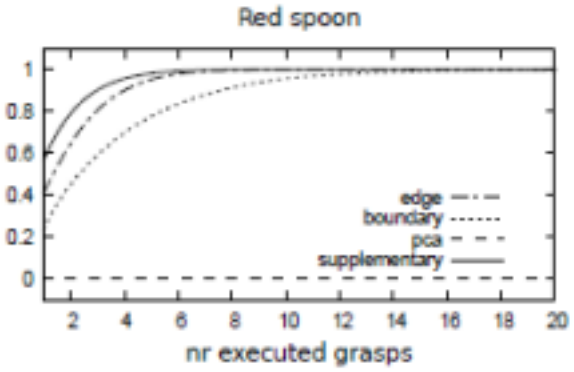


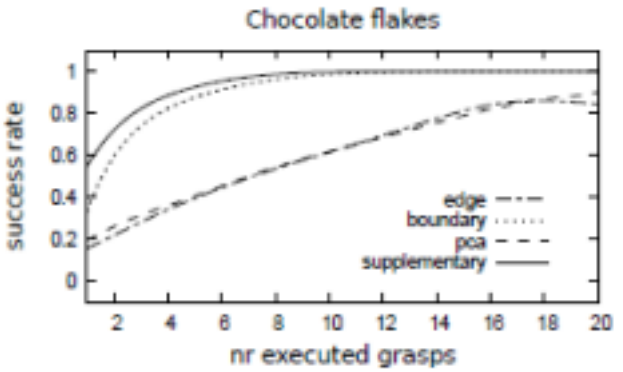
N. Pugeault, F. Wörgötter and N. Krüger. Visual Primitives: Local, Condensed, and Semantically Rich Visual Descriptors and their Applications in Robotics. International Journal of Humanoid Robotics, Volume: 7, Issue 3, pp. 379-405, 2010.

Grasp Affordances of Edges and Surfaces





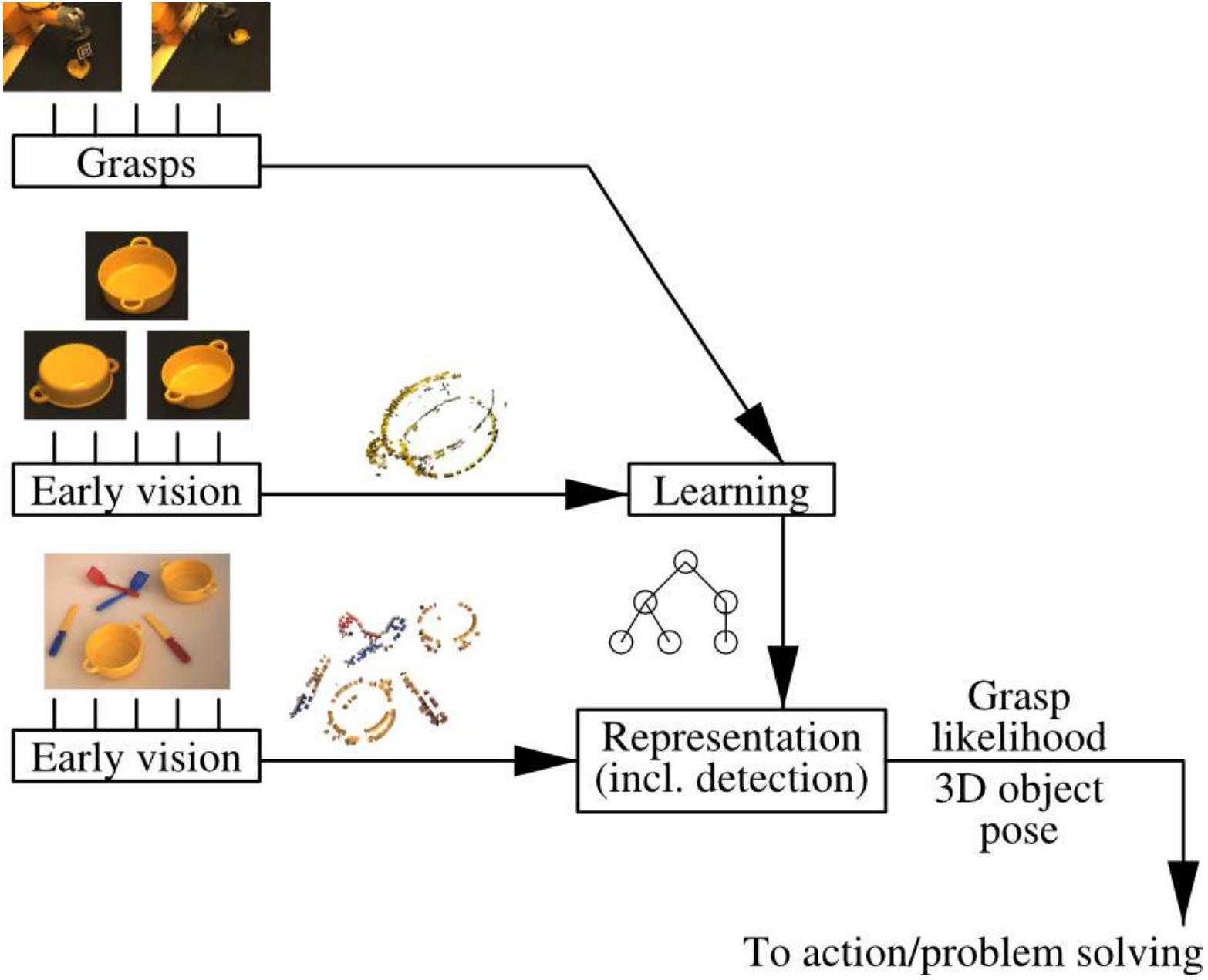




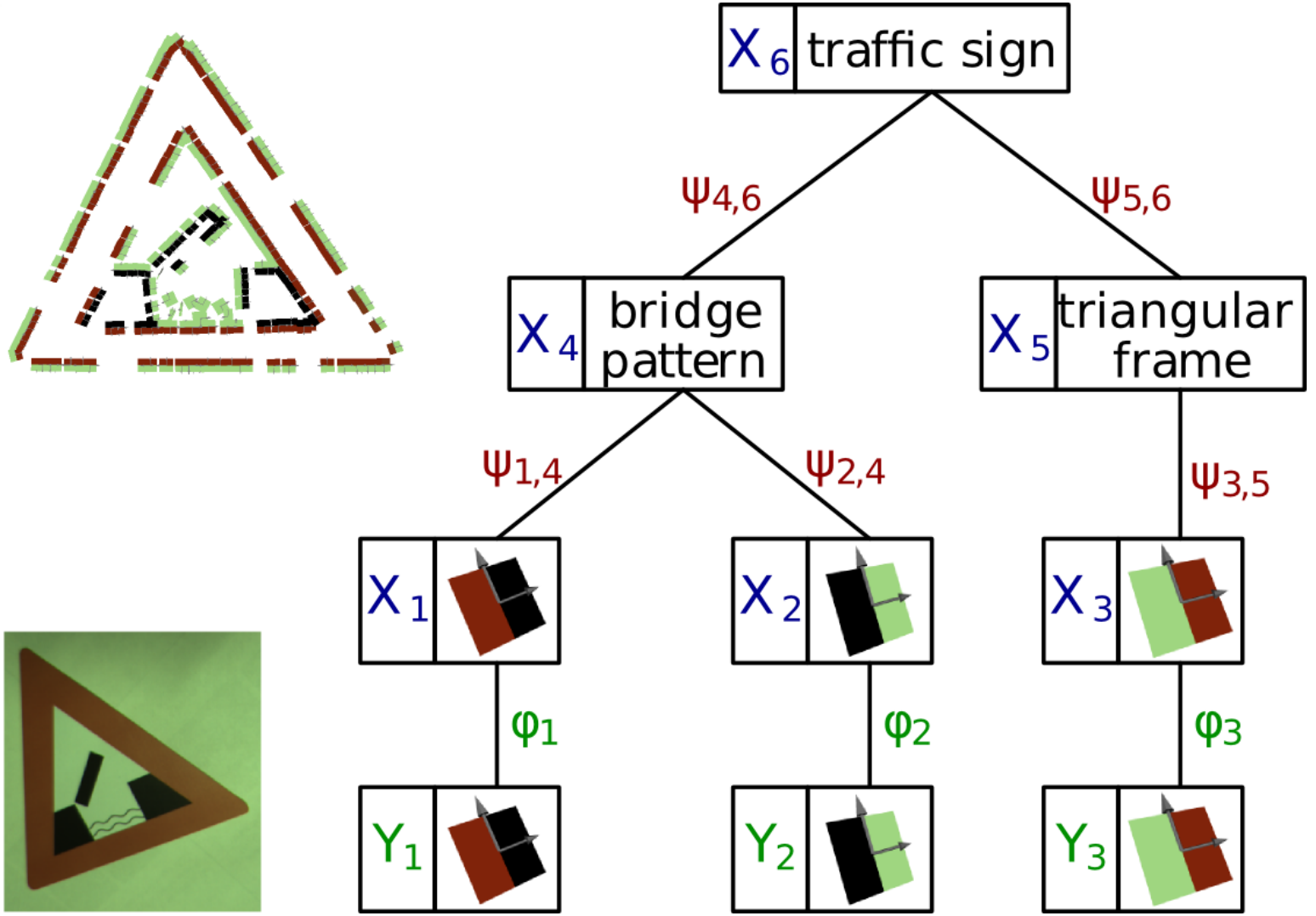
M. Popović, G. Kootstra, J. A. Jørgensen, D. Kragic and N. Krüger. Grasping Unknown Objects using an Early Cognitive Vision System for General Scene Understanding. IROS 2011 (nominated as one of the finalists for an IROS award)

G. Kootstra, M. Popovic, J. A. Jorgensen, K. Kuklinski, K. Miatliuk, D. Kragic and N. Krüger. Enabling grasping of unknown objects through a synergistic use of edge and surface information. International Journal of Robotics Research, vol. 31, no. 10, pp. 1190 - 1213, 2012.

Hierarchical Markov Models

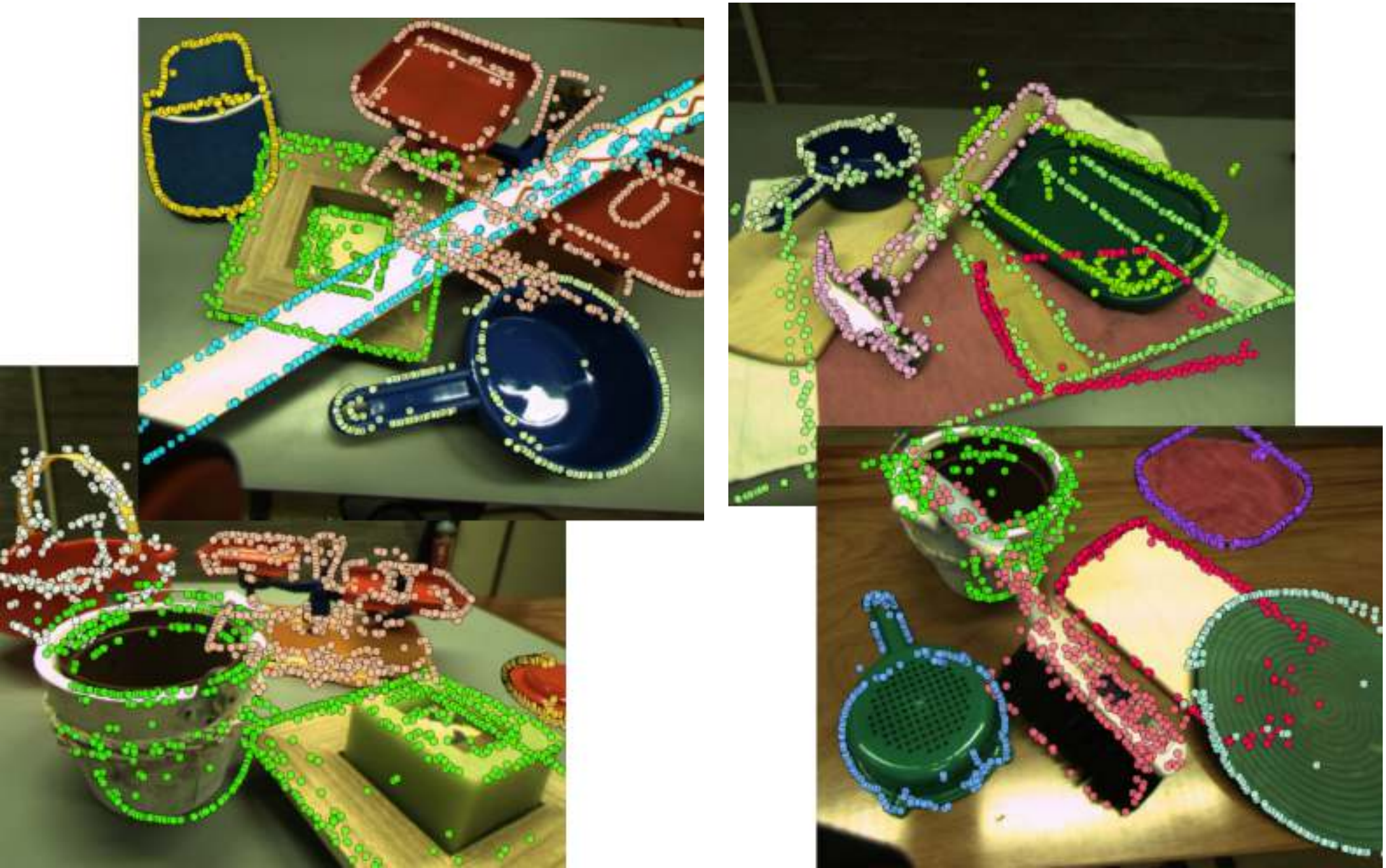


Hierarchical Markov Models

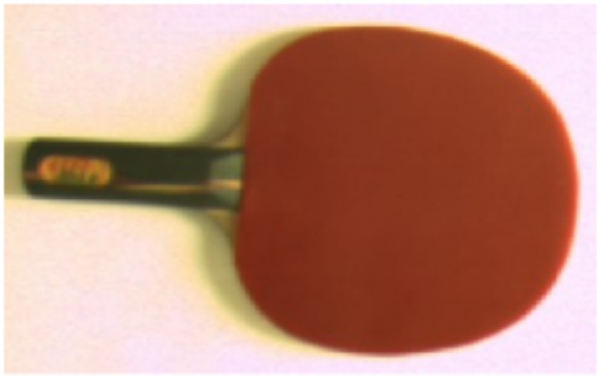


Renaud Detry, Nicolas Pugeault, Justus Piater, A Probabilistic Framework for 3D Visual Object Representation. PAMI 31 (10), pp. 1790–1803, 2009

Hierarchical Markov Models

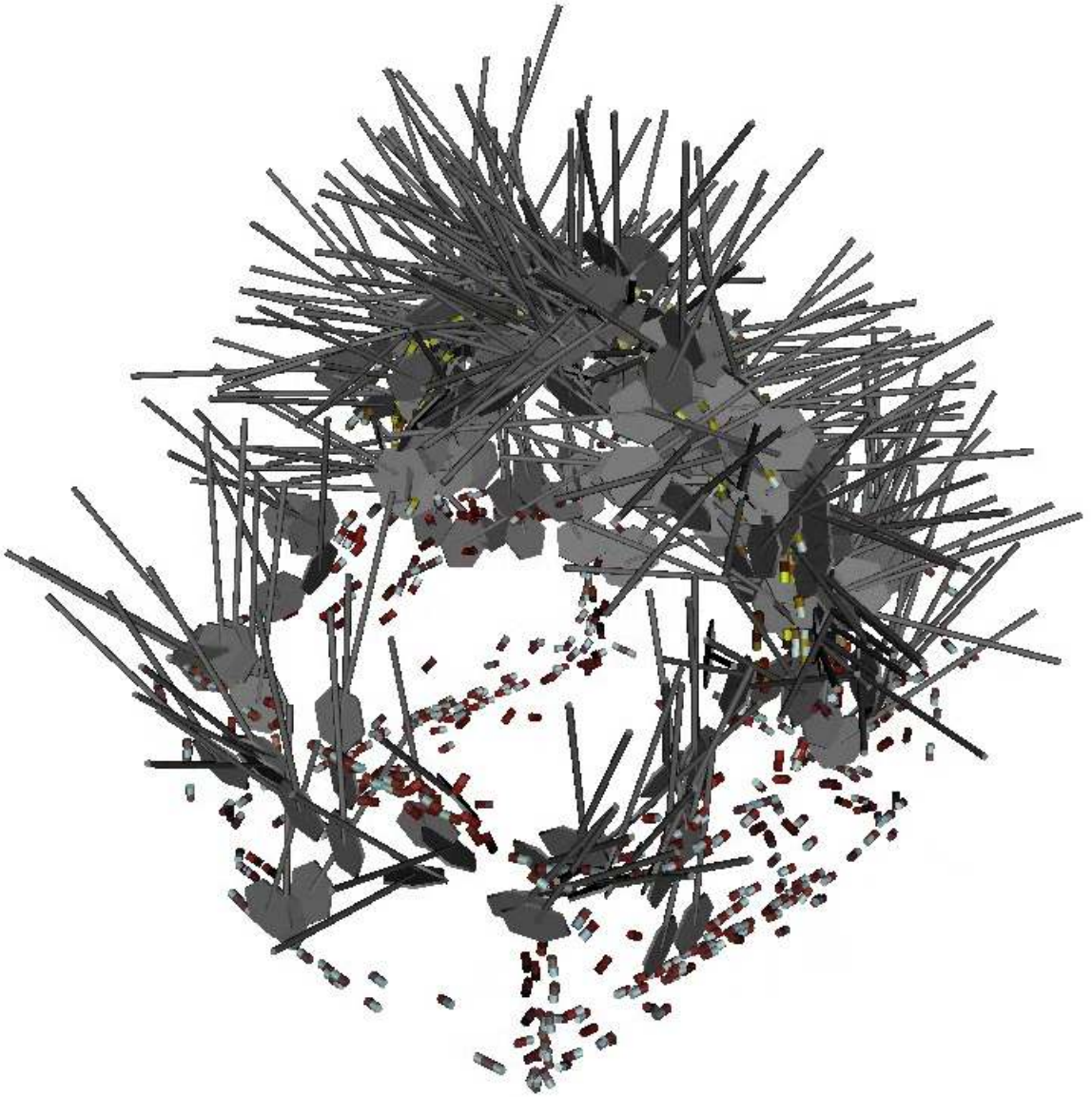


Grasp Densities



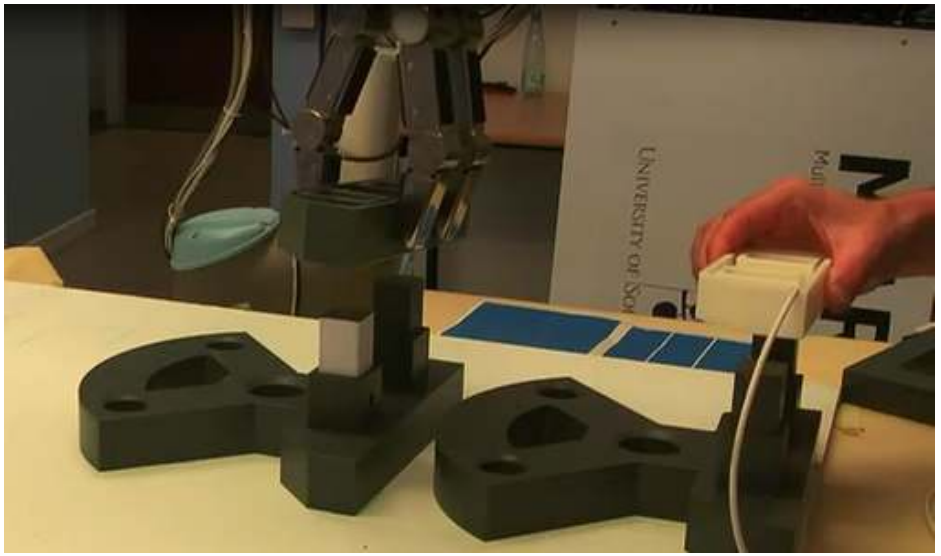
R. Detry, D. Kraft, O. Kroemer, L. Bodenhagen, J. Peters, N. Krüger, J. Piater, Learning Grasp Affordance Densities. Paladyn Journal of Behavioral Robotics 2 (1), pp. 1–17, 2011

A Grasp Density




STAUBLI

Peg-in-Hole with Force Feedback

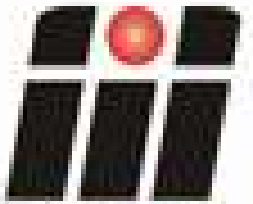


- Movement and force/torque trajectories are captured during human demonstration.
- Learn variable stiffness by policy-search RL.
- Iterative, on-line adaptation: Positional trajectories are adapted to match the demonstrated force/torque profile.

B. Nemec, F. Abu-Dakka, J. A. Jørgensen, T. R. Savarimuthu, B. Ridge, H. G. Petersen, J. Jouffroy, N. Krüger, and A. Ude, Transfer of Assembly Operations to New Workpiece Poses by Adaptation to the Desired Force Profile, ICAR 2013







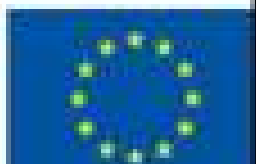
IntellAct

DMP - Learning

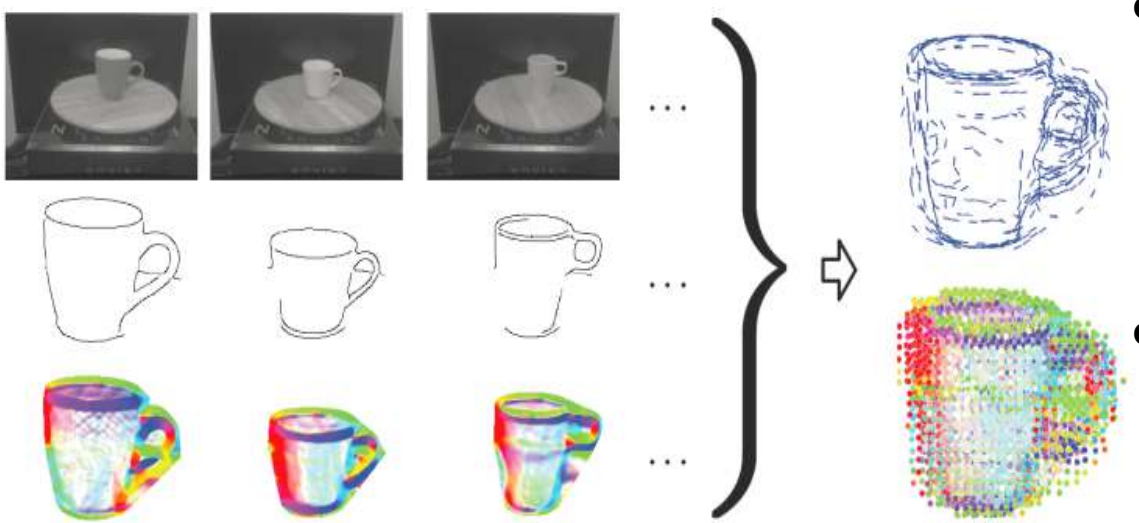
Thiusius R. Savarimuthu, Anders G. Buch, Wail Mustafa, Yang Yang, Aljaz Kramberger, Bojan Nemec

IntellAct (2011-2014): Intelligent Observation and Execution of Actions and Manipulations



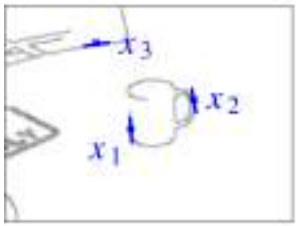


3D Inference From View-based Models

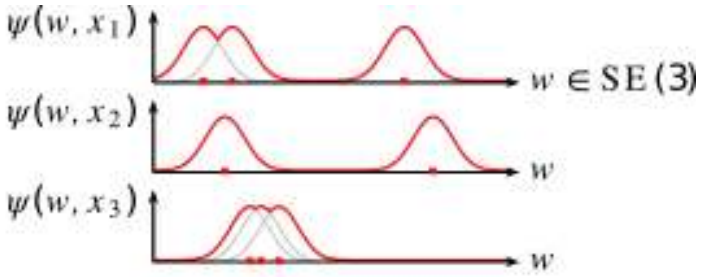


- 2D edge and gradient distributions
- probabilistic 6-DoF pose inference

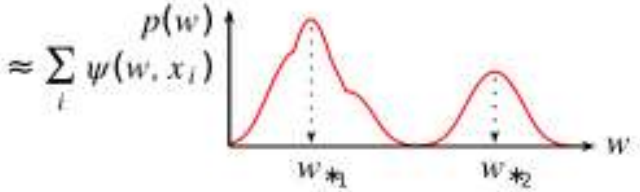
(a)



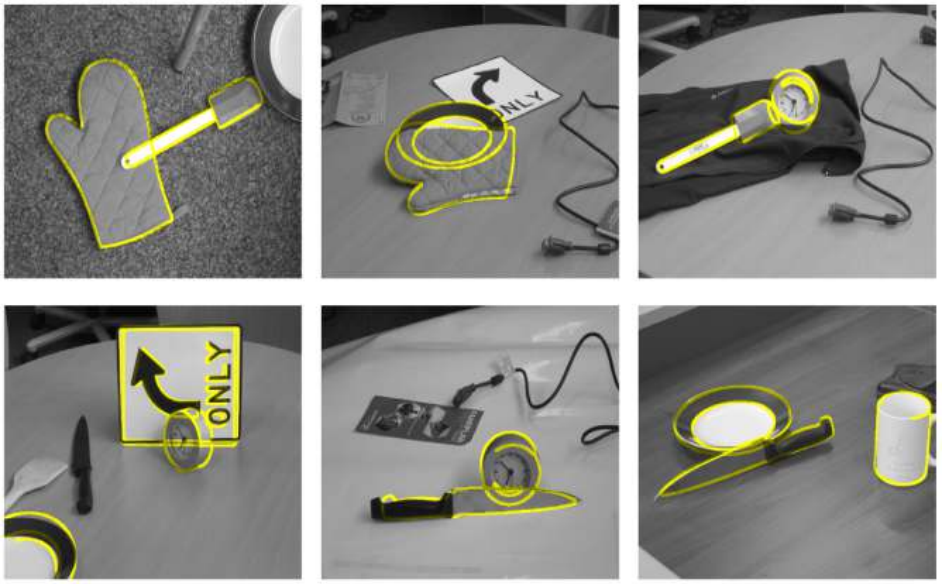
(b)



(c)

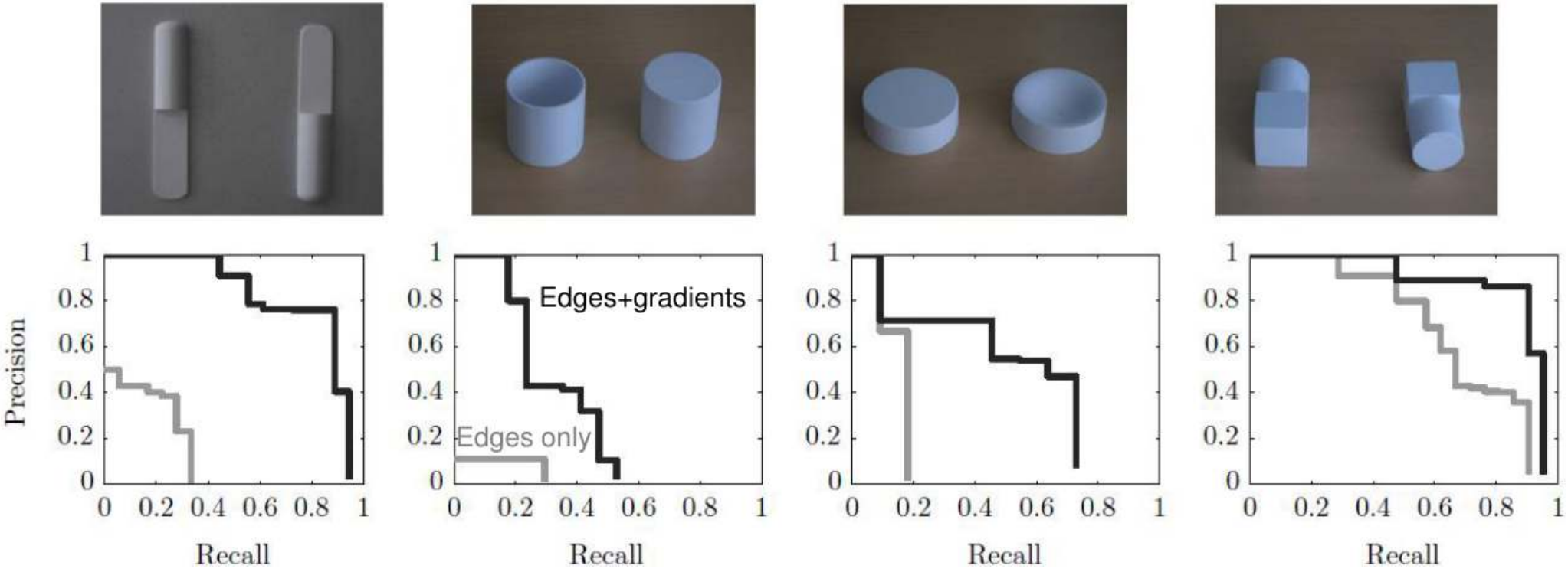


(d)

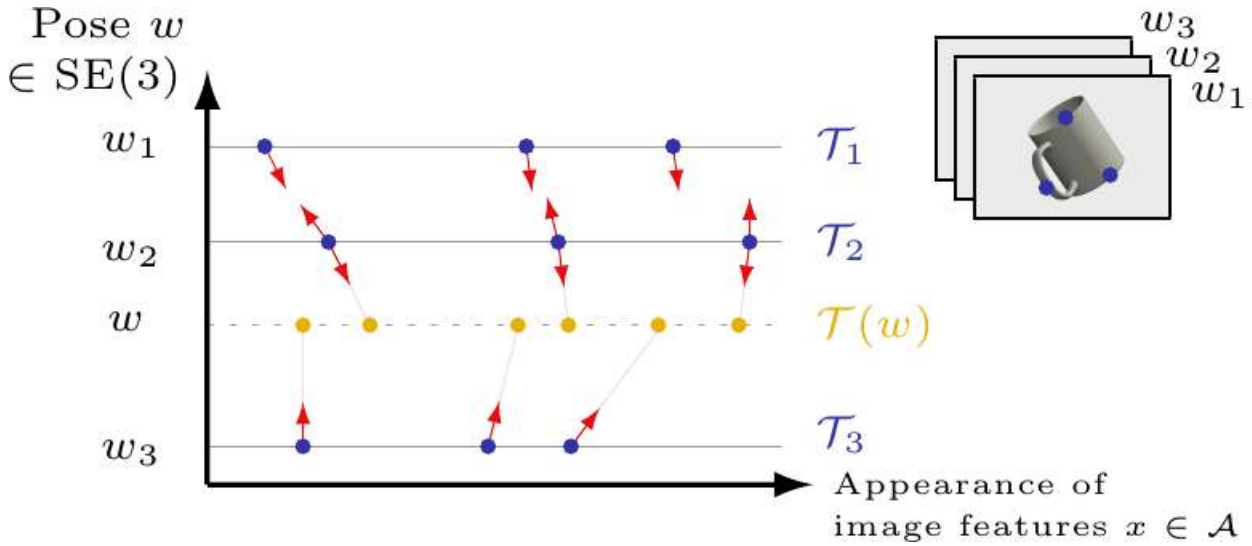


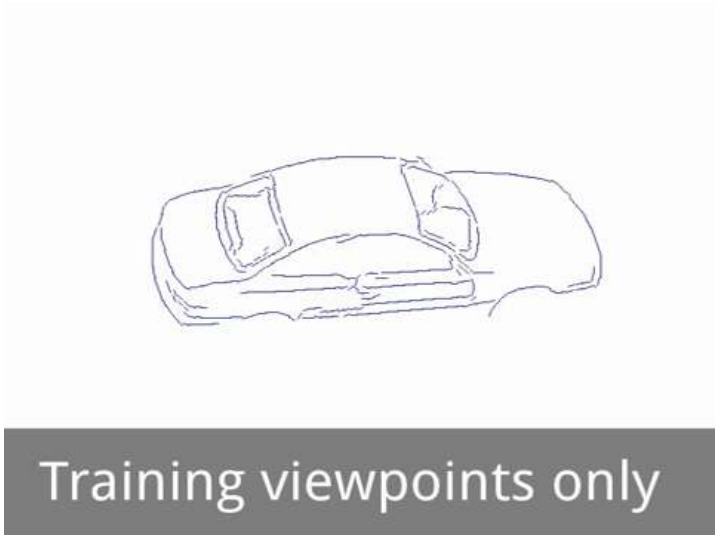
Damien Teney, Justus Piater, Multiview feature distributions for object detection and continuous pose estimation. Computer Vision and Image Understanding 125, pp. 265–282, 2014

Pose Inference from Shading



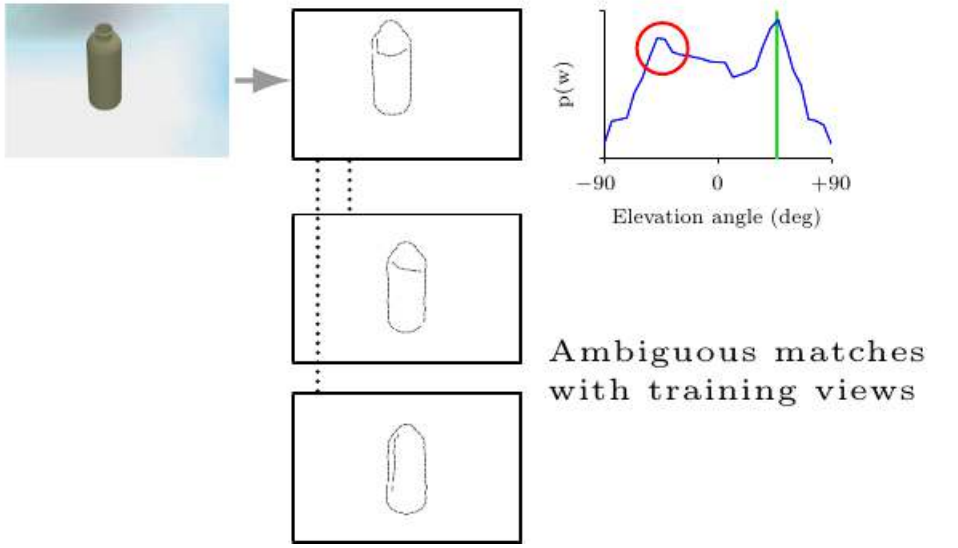
View Interpolation and Parallax





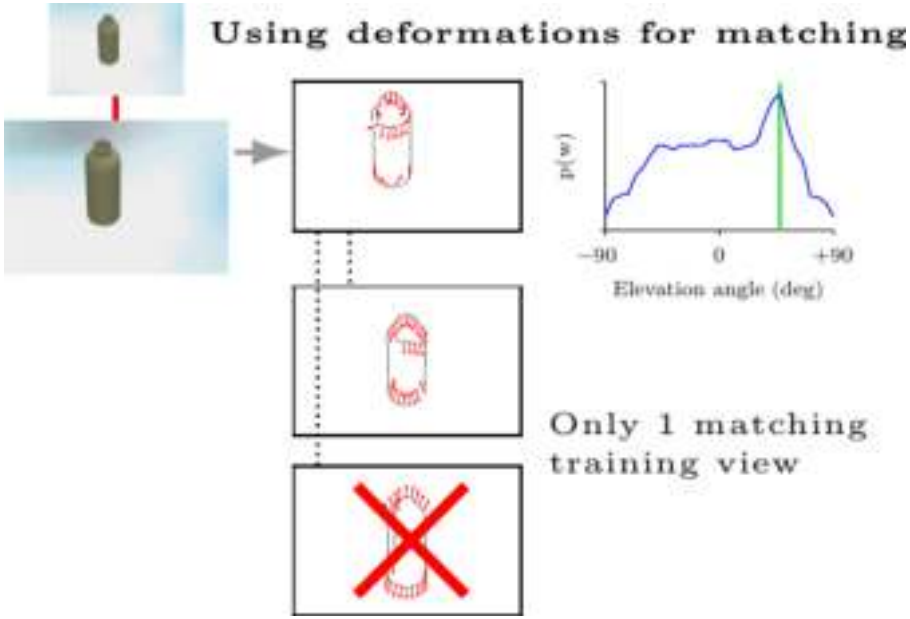
Training viewpoints only

Without detecting deformations in test scene



Ambiguous matches
with training views

Using deformations for matching



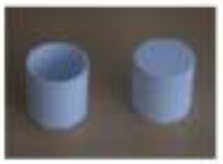
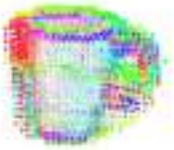
Only 1 matching
training view

Outline

- We've seen:
  - Structure
    - convex parts
    - ECV feature hierarchy
  - Association of
    - structural models and 6-DoF pose ECV observations
    - grasps to structural models
    - wrenches to positional trajectories
    - models and 6-DoF pose to 2D observations
    - pose cues to shading
    - pose cues to parallax
- What next?



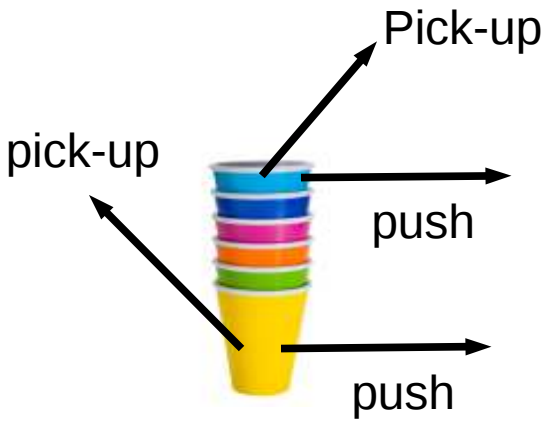


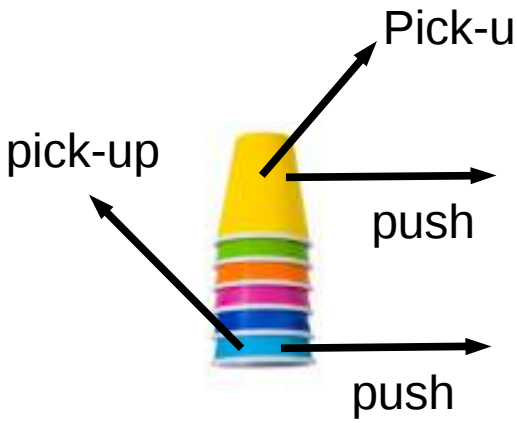





Syntax

- Single-digit counting/incrementation:
  - memorized sequence = associated successors
- Carrying to new column (9 + 1 = 10):
  - syntactic rule
- Addition of single-digit numbers:
  - reduced to repeated incrementation
  - memorized = sums associated to pairs of numbers
- Addition of general numbers:
  - alternating associative and syntactic steps
- Multiplication of single-digit numbers:
  - reduced to repeated addition
  - memorized = products associated to pairs of numbers
- Multiplication of general numbers:
  - alternating associative and syntactic steps

Multi-Object Concepts



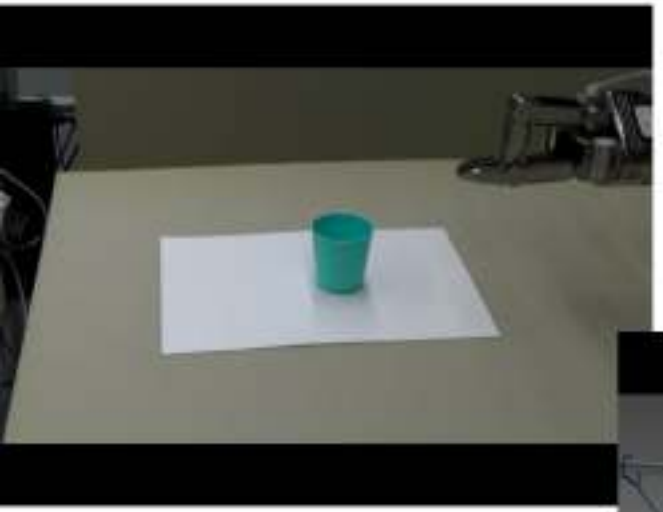


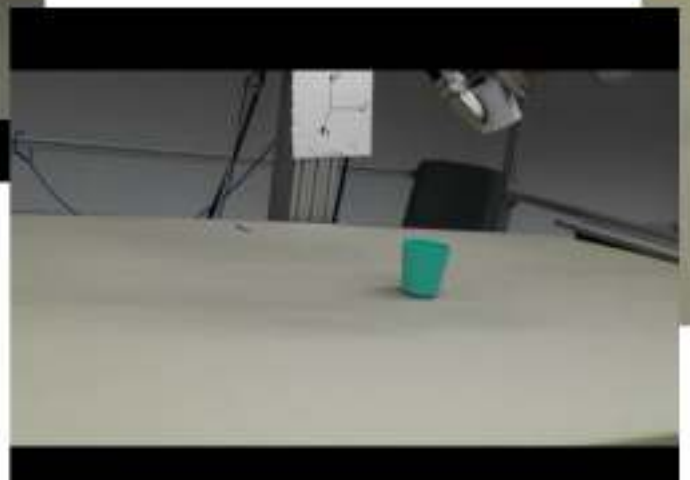
Pick-up

Structural Bootstrapping

Stacking Learning Problems

- Shape and size features from Kinect point clouds
- Actions: side-poke, top-poke, front-poke, stack
- Poke-effects: pushed, rolled, toppled, resisted, nothing
- Stack-effects: piled-up, inserted-in, covered, tumble-over

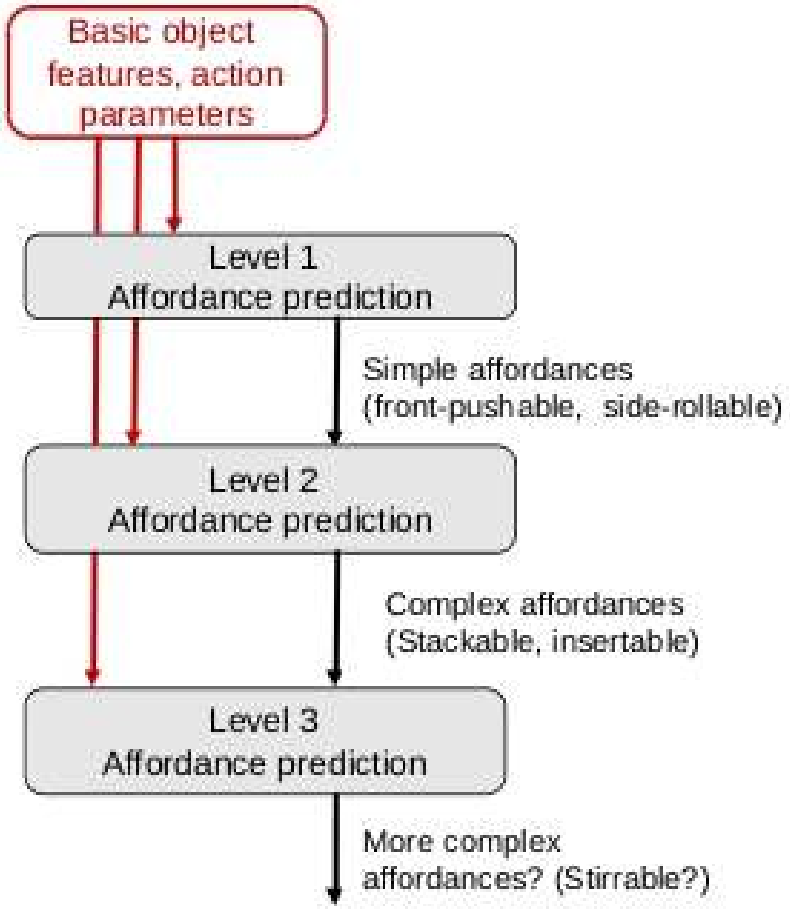


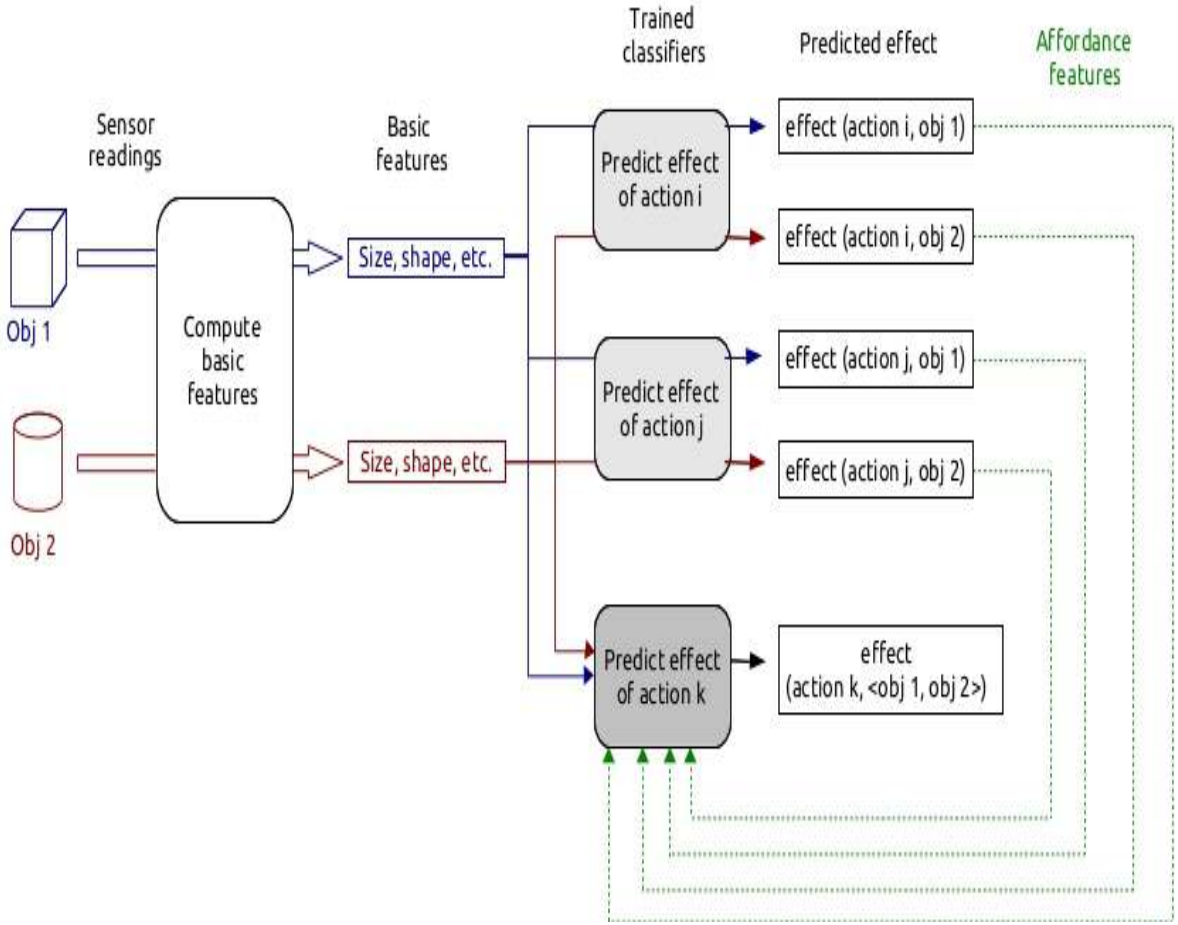


Emre Ugur et al., submitted

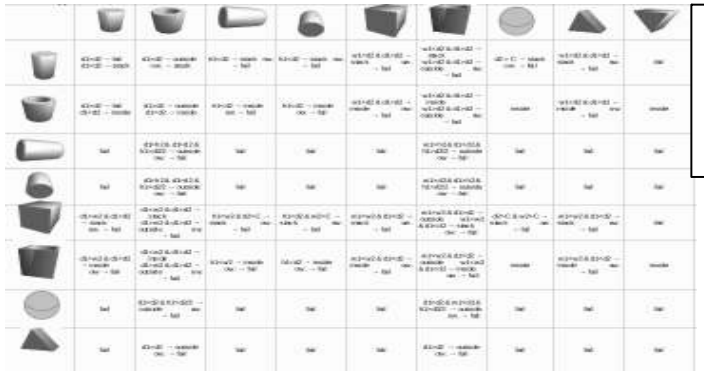
Stacking Learning Problems

- Encode structural similarities within affordances
- Bootstrap paired-object affordance learning with learned single-object affordances





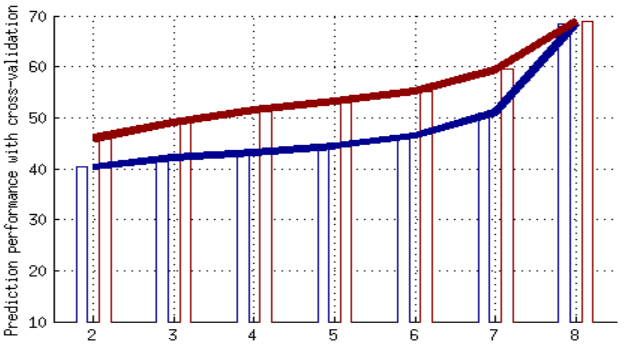
Stacking Learning Problems

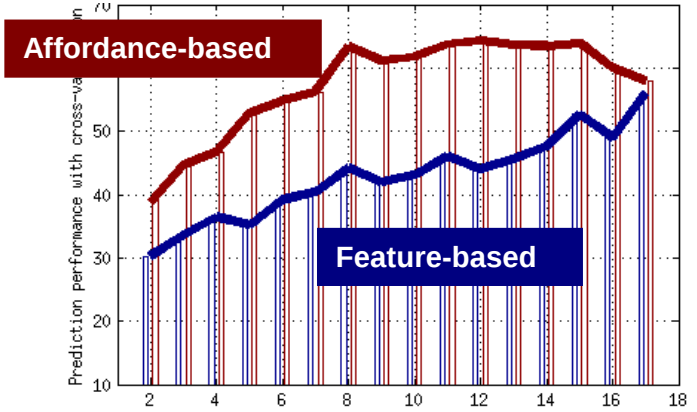


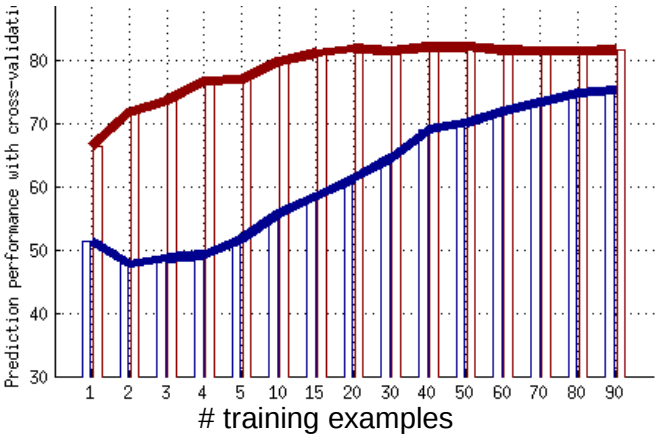
9 categories
Rule based effects
Hand-coded shape features

20 interactions,
Observed effects
Kinect based shape features

83 objects
Manually coded 83x83 effects
Kinect based shape features







Discussion

- Should robots operate associatively?
  - Computers are great at symbolic reasoning! Let them play to their strengths! Why imitate humans?
- Perception is poor at providing reasonable symbols!

  (Perhaps this is why humans did not evolve great facility at symbolic reasoning?)
- Machine Learning is strong.
  - Let's put it to use in conjunction with structure and syntax!Let's get robots to thin-slice!
- Use numerous, complementary perceptual modalities.
  - vision, wrench, touch, proprioception, …
- Simplify perception and control by exploiting the interaction dynamics.
  - synergies, soft hands, external constraints, …